Communications of the Korean Statistical Society
2009, Vol. 16, No. 1, 169–183

Unrelated Question Model in Sensitive Multi-Character Surveys

Sukhjinder Singh Sidhu[a], Mohan Lal Bansal[a], Jong-Min Kim[1,b], Sarjinder Singh[c]

[a]Dept.of Mathematics, Statistics and Physics, Punjab Agricultural Univ.,
[b]Div. of Science and Mathematics, Univ. of Minnesota, [c]Dept. of Mathematics, Texas A&M Univ.

Abstract

The simplicity and wide application of Greenberg *et al.* (1971) prompts to propose a set of alternative estimators of population total for multi-character surveys that elicit simultaneous information on many sensitive study variables. The proposed estimators take into account the already known rough value of the correlation coefficient between Y(the characteristic under study) and p(the measure of size). These estimators are biased, but it is expected that the extent of bias will be smaller, since the proposed estimators are suitable for situations in between those optimum for the usual estimators and the estimators based on multi-characters for no correlation. The relative efficiency of the proposed estimators has been studied under a super population model through empirical study. It has been found through simulation study that a choice of an unrelated variable in the Greenberg *et al.* (1971) model could be made based on its correlation with the auxiliary variable used at estimation stage in multi-character surveys.

Keywords: Total estimation, RRT, sensitive multi-characteristics, mean square error, super population model, cost aspects and empirical study.

1. Introduction

The well-known Hansen and Hurwitz (1943) estimator of population total for probability proportional to size and with replacement sampling (PPSWR) is given by

$$\widehat{Y}_{\mathrm{HH}} = \frac{1}{n}\sum_{i=1}^{n}\frac{y_i}{p_i}, \quad \text{where } p_i = x_i\left(\sum_{i=1}^{N} x_i\right)^{-1}. \tag{1.1}$$

In sample surveys of many variables, some of the study variables may be poorly correlated with the selection probabilities. In this the use of usual estimators available in literature results in larger variance. Rao (1966) has provided alternative estimators when the study variable and size measure are unrelated and demonstrated that these alternative estimators are more efficient though biased. But Rao's (1966) model is not commonly encountered in practice since the correlation is not always zero. Bansal and Singh (1985) developed a transformed estimator of population total suitable for the characteristics covering entire range of positive correlation. Amahia *et al.* (1989) suggested simple alternatives to the transformations in Bansal and Singh (1985). The transformations of selection probabilities used are as follows:

[1] Corresponding author: Associate Professor, Statistics Discipline, Division of Science and Mathematics, University of Minnesota, Morris, MN 56267-2134, USA. E-mail: jongmink@morris.umn.edu

$$p_{i0}^* = \frac{1}{N}, \qquad \text{[Rao (1966)]} \tag{1.2}$$

$$p_{i1}^* = \left(1 + \frac{1}{N}\right)^{1-\rho} (1 + p_i)^\rho - 1, \qquad \text{[Bansal and Singh (1985)]} \tag{1.3}$$

$$p_{i2}^* = \frac{1}{N}(1 - \rho) + \rho p_i, \qquad \text{[Amahia et al. (1989)]} \tag{1.4}$$

$$p_{i3}^* = \left(\frac{1}{N}\right)^{1-\rho} p_i^\rho, \qquad \text{[Amahia et al. (1989)]} \tag{1.5}$$

$$p_{i4}^* = \left\{N(1 - \rho) + \frac{\rho}{p_i}\right\}^{-1}, \qquad \text{[Amahia et al. (1989)]} \tag{1.6}$$

$$p_{i5}^* = \frac{1}{N}\left(1 - \rho^{\frac{1}{3}}\right) + \rho^{\frac{1}{3}} p_i, \qquad \text{[Grewal et al. (1997)].} \tag{1.7}$$

On the basis of these transformations, following types of estimators of population total Y under PPSWR sampling are available in the literature:

$$\left(\widehat{Y}_{\text{pps}}\right)_h = \frac{1}{n} \sum_{i=1}^{n} \frac{y_i}{p_{ih}^*}, \quad h = 1, 2, 3, 4, 5. \tag{1.8}$$

The transformations p_{ih}^* $(h = 1, 2, 3, 4, 5)$ at (1.3) to (1.7) of the selection probabilities p_i are useful for positive correlation between y_i and p_i variables, whereas transformation (1.2) is useful under no correlation situation. Interestingly for $\rho = 0$, p_{ih}^* $(h = 1, 2, 3, 4, 5)$ reduce to p_{i0}^* at (1.2) and for $\rho = 1$ these transformations reduce to original selection probabilities p_i. For detail one can refer to Arnab (2001) and Singh (2003).

The surveys on human population had established the fact that the direct question about sensitive characters often result in either refusal to respond or falsification of the answer. This can bias the estimates. Warner (1965) developed an interviewing procedure designed to reduce or eliminate this bias and called it as Randomized Response Technique(RRT). It is beneficial to combine multi-characteristics and RRT. Bansal *et al.* (1994) and Grewal *et al.*(1997) had discussed the multi-characteristics in RRT to estimate population total.

It was felt that the confidence of the respondents in anonymity provided by RRT and hence reliability of their responses, might be further enhanced if one of the two question belong to non sensitive, innocuous attribute unrelated to the sensitive characteristics. Greenberg *et al.* (1971) developed the work for quantitative responses and found that his unrelated question technique was more efficient than the Warner (1965) model.

2. UQ Model

In the quantitative unrelated question (UQ) random response model, using two questions, the overall distribution of responses is comprised of numerical answers to both questions, the answers being indistinguishable as to question. This distribution is a mixture of two pure distributions, which must be statistically separated to provide meaningful estimates of the parameters of interest. The population means of both the sensitive(Y) and unrelated non-sensitive(U) variables are μ_y and μ_U with their respective variances σ_y^2 and σ_U^2.

When the value of U(total of unrelated character) is known in advance we select one sample of size n. The respondent in the sample is provided with a randomization device, with probability T and $(1-T)$, respectively, consisting of sensitive and non sensitive statements:

(i) About how much money in dollars did the head of household, earn last year?

(ii) About how much average money in dollars do you think the head of a household of your size earns in a year?

The respondent selects randomly one of the two statements, unobserved by the interviewer, and reports the answer. Let response from i^{th} individual in the sample for the characteristic under study be denoted by r_i.

$$r_i = T\ y_i + (1-T)\mu_i \tag{2.1}$$

with

$$V(r_i) = T(1-T)(y_i-\mu_i)^2 . \tag{2.2}$$

Keeping in view the importance of this model, we extend the method to multi-character surveys to propose estimators of population total. The behavior of the proposed estimators has been examined under the super population model given below.

3. Super Population Model

A general super population model for sensitive characteristic under study is:

$$Y_i = \beta p_i + e_i, \quad i = 1, 2, \ldots, N, \tag{3.1}$$

where e_i's are the error terms such that:

$$E_m(e_i \mid p_i) = 0, \tag{3.2}$$

$$E_m\left(e_i e_j \mid p_i p_j\right) = 0 \tag{3.3}$$

and

$$E_m\left(e_i^2 \mid p_i\right) = a p_i^g, \quad a > 0,\ g \geq 0. \tag{3.4}$$

Here $E_m(e_i^2 \mid p_i)$ is the residual variances of Y for given p_i. The expected value of this residual variance in the super population model is given by:

$$E_m\left(a p_i^g\right) = a E_m(p_i^g) \tag{3.5}$$

and when the infinite super population is simulated by a finite large population of N units having the same characteristics it will be reduced to:

$$E\left(a p_i^g\right) = \frac{a}{N}\sum_{i=1}^{N} p_i^g . \tag{3.6}$$

Also the expected value of residual variance is known to be given by $\sigma_y^2\left(1-\rho^2\right)$.

Thus we have:

$$\frac{a}{N}\sum_{i=1}^{N} p_i^g = \sigma_y^2\left(1-\rho^2\right) \tag{3.7}$$

or

$$\sigma_y^2 = \frac{\frac{a}{N}\sum_{i=1}^{N} p_i^g}{1-\rho^2}. \tag{3.8}$$

The value of the regression coefficient is given by:

$$\beta^2 = \rho^2 \frac{\sigma_y^2}{\sigma_p^2} = \frac{\rho^2}{1-\rho^2}\left(\frac{a}{N}\sum_{i=1}^{N}\frac{p_i^g}{\sigma_p^2}\right), \tag{3.9}$$

where

$$\sigma_p^2 = \frac{1}{N}\left(\sum_{i=1}^{N} p_i^2 - \frac{\left(\sum_{i=1}^{N} p_i\right)^2}{N}\right). \tag{3.10}$$

The super population model for unrelated non-sensitive question is:

$$U_i = \beta^* p_i + e_i^*, \quad i = 1, 2, \ldots, N, \tag{3.11}$$

where e_i^*'s are the error terms satisfying all the conditions at (3.2), (3.3) and (3.4).

It is assumed for simplicity that means of Y_i and U_i are different but the residual variances of U for $p = p_i$, *i.e.* $E\left(e_i^{*2} \mid p_i\right)$ is same as of Y.

Similarly

$$\beta^{*2} = \rho^{*2} \frac{\sigma_U^2}{\sigma_p^2} = \frac{\rho^{*2}}{1-\rho^{*2}}\left(\frac{a}{N}\sum_{i=1}^{N}\frac{p_i^g}{\sigma_p^2}\right). \tag{3.12}$$

We first obtain the estimator of population total for PPSWR.

4. Estimator $(\widehat{Y}_1)$

When the value of U(total of unrelated character) is known in advance we select one sample of size n. The respondents in the sample are provided with a randomization device consisting of sensitive and non sensitive statements with probability T and $(1-T)$ respectively. The respondent selects randomly one of the two statements, unobserved by the interviewer and reports the answer. Let response from i^{th} individual in the sample for the characteristic under study be denoted by r_i.

$$r_i = Ty_i + (1-T)\mu_i. \tag{4.1}$$

The estimator of population total $(\widehat{Y}_1)$ for PPSWR is obtained as

$$\widehat{Y}_1 = \frac{1}{T}\left\{\frac{1}{n}\sum_{i=1}^{n}\frac{r_i}{p_i} - (1-T)\,U\right\}. \tag{4.2}$$

The variance of the randomized response of i^{th} individual is

$$V(r_i) = T(1-T)(y_i - \mu_i)^2. \tag{4.3}$$

Using this result it can be shown easily that the estimator $(\widehat{Y}_1)$ is unbiased. The variance of the estimator $(\widehat{Y}_1)$ is given in the following theorem.

Theorem 1. *The variance of the estimator* $(\widehat{Y}_1)$ *given in* (4.2) *is given by*

$$V\left(\widehat{Y}_1\right) = \frac{1-T}{nT}\sum_{i=1}^{N}\frac{(Y_i - U_i)^2}{p_i} + \frac{1}{nT^2}\sum_{i=1}^{N} p_i\left\{T\left(\frac{Y_i}{p_i} - Y\right) + (1-T)\left(\frac{U_i}{p_i} - U\right)\right\}^2. \tag{4.4}$$

Proof: Please see the Appendix A.

We now extend the theory for the estimator obtained above to propose the estimators of population total in case of multi-character surveys. □

5. Proposed Estimator $(\widehat{Y}_2)$

The proposed estimators of population total $(\widehat{Y}_2)_h$ for multi-characteristics are given by

$$\left(\widehat{Y}_2\right)_h = \frac{1}{T}\left\{\frac{1}{n}\sum_{i=1}^{n}\frac{r_i}{p_{ih}^*} - (1-T)\,U\right\}, \tag{5.1}$$

where p_{ih}^* are defined in (1.2) to (1.7). The proposed estimators are biased and the bias in the the estimators $(\widehat{Y}_2)_h$ is given by

$$B\left(\widehat{Y}_2\right)_h = \sum_{i=1}^{N}\left(\frac{p_i}{p_{ih}^*} - 1\right)\left\{Y_i + \frac{(1-T)}{T}U_i\right\}. \tag{5.2}$$

One can easily see that the variance of the estimator $(\widehat{Y}_2)_h$ is given by

$$V\left(\widehat{Y}_2\right)_h = \frac{1-T}{nT}\sum_{i=1}^{N}\frac{(Y_i - U_i)^2\, p_i}{p_{ih}^{*2}} + \frac{1}{nT^2}\left[\sum_{i=1}^{N}\frac{(TY_i + (1-T)\,U_i)^2\, p_i}{p_{ih}^{*2}} - \left\{\sum_{i=1}^{N}\frac{(TY_i + (1-T)\,U_i)\, p_i}{p_{ih}^{*}}\right\}^2\right]. \tag{5.3}$$

Please see the full derivation of the equation (5.3) in the Appendix A.

To obtain the expected Mean Square Error (MSE) of proposed estimators $(\widehat{Y}_2)_h$ under super population model we have the following theorem.

Theorem 2. *The expected value of MSE of* $(\widehat{Y}_2)_h$ *under the superpopulation model is*

$$E_m\left[MSE\left(\widehat{Y}_2\right)_h\right] = \frac{1}{n}A_1 + A_2, \tag{5.4}$$

where

$$A_1 = \left[\left(\frac{1-T}{T}\right)\left((\beta - \beta^*)^2\sum_{i=1}^{N}\frac{p_i^3}{p_{ih}^{*2}} + 2a\sum_{i=1}^{N}\frac{p_i^{g+1}}{p_{ih}^{*2}}\right) + \frac{1}{T^2}\left\{a\left(T^2 + (1-T)^2\right)\left(\sum_{i=1}^{N}\frac{p_i^{g+1}}{p_{ih}^{*2}} - \sum_{i=1}^{N}\frac{p_i^{g+2}}{p_{ih}^{*2}}\right) + (\beta T + (1-T)\beta^*)^2\left(\sum_{i=1}^{N}\frac{p_i^3}{p_{ih}^{*2}} - \left(\sum_{i=1}^{N}\frac{p_i^2}{p_{ih}^{*}}\right)^2\right)\right\}\right] \tag{5.5}$$

and

$$A_2 = \left\{\left(\sum_{i=1}^{N} \frac{p_i^2}{p_{ih}^*} - 1\right)\left(\beta + \frac{1-T}{T}\beta^*\right)\right\}^2 + a\left(1 + \frac{(1-T)^2}{T^2}\right)\sum_{i=1}^{N}\left(\frac{p_i^{g+2}}{p_{ih}^{*2}} + p_i^g - 2\frac{p_i^{g+1}}{p_{ih}^*}\right). \tag{5.6}$$

Proof: Please see the Appendix A.

When the value of non-sensitive question is known in advance. We choose the strategy, which for a fixed cost can estimate Y with maximum accuracy. For this we find the minimum expected mean square error for fixed cost under super population model. This we do in the following theorem. □

Theorem 3. *Under superpopulation model, for the fixed cost C_0 the minimum expected mean square error of estimator $(\widehat{Y}_2)_h$ is given by*

$$E_m\left[MSE\left(\widehat{Y}_2\right)_h\right] = \frac{C_1}{C_0}A_1 + A_2, \tag{5.7}$$

where A_1 and A_2 are defined in (5.5) *and* (5.6) *and C_1 cost of processing per unit in the sample.*

Proof: Please see the Appendix A. □

6. Empirical Study

To investigate into the performance of the proposed estimators we resort to an empirical study under super population model given in Section 3. For this the relative efficiency under unrelated question model(RE_*h*) of the proposed estimators $(\widehat{Y}_2)_h$ for $h = 1, 2, 3, 4, 5$ with respect to $(\widehat{Y}_2)_0$ is given by

$$(\text{RE}\_h) = \frac{E_m\left[\text{MSE}\left(\widehat{Y}_2\right)_0\right]}{E_m\left[\text{MSE}\left(\widehat{Y}_2\right)_h\right]} \times 100, \tag{6.1}$$

where symbols have their usual meanings. The probability associated with the statements in the device is 0.7 and 0.3, respectively. The choice of $T = 0.7$ in the Greenberg *et al.* (1971) seems to be a reasonable choice, because a very high value of T may effect the respondents' coopeartion while asking a question through the randomization device. The density functions for the auxiliary variable, which is assumed to have correlation of values ρ in the range 0 to 1 with the study variables, are presented in Table 1.

For the sensitive character value of correlation coefficient between X and Y is $\rho = 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9$, but for unrelated question, correlation coefficient $\rho^* = 0.15$, 0.65, 0.95 is used. Note that ρ^* is the value of correlation coefficient between the unrelated character variable and the auxiliary variable or say the selection probabilities p_i. Thus, it could always be feasible to select an unrelated character variable which may have either low, moderate or high value of correlation coefficient with the selection probabilities p, thus we considered only three such values of ρ^*. As the nature of the sensitive variables y_is remains unpredictable, thus we decided to consider entire range of values of the correlation coefficient ρ between 0 and 1. A PPSWR sample of size 20 is considered as drawn from a population consisting of 100 respondents. The computations are given in Appendix B. The results obtained from these computations are given below in Table 2.

From this table it is clear that the proposed estimators fare better than the usual estimator for all the p_{ih}^* $(h = 1, 2, 3, 4, 5)$ in majority of the cases. It is to pointed out that if the value of ρ^* is low

Table 1: Density functions for various probability distributions

Sr. No.	Distribution	Density function	Range
1	Right Triangular	$f(x) = 2(1-x)$	$0 \le x \le 1$
2	Exponential	$f(x) = e^{-x}$	$x \ge 0$
3	Chi-Square at $\nu = 6$	$f(x) = \frac{1}{2^{\frac{\nu}{2}}\Gamma_{\frac{\nu}{2}}} e^{-\frac{x}{2}} x^{\frac{\nu-2}{2}}$	$x \ge 0$
4	Gamma, $\alpha = 2$, $\beta = 1$	$f(x) = \frac{1}{\beta^{\alpha}\Gamma_{\alpha}} x^{\alpha-1} e^{-\frac{x}{\beta}}$	$x \ge 0$
5	Normal	$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$	$-\infty < x < +\infty$
6	Log Normal	$f(x) = \frac{1}{x\sqrt{2\pi}} e^{-\frac{1}{2}(\log(x))^2}$	$x > 0$
7	Beta, $\alpha = 3$, $\beta = 2$	$f(x) = \frac{1}{B(\alpha,\beta)} x^{\alpha-1}(1-x)^{\beta-1}$	$0 \le x \le 1$

Table 2: Correlation range for which the proposed estimator $(\widehat{Y}_2)_h$ of population total of a sensitive character is more efficient than usual estimator for p^*_{ih} given in (1.2) to (1.7).

Distribution	$g = 0$			$g = 1$			$g = 2$		
ρ^*	0.15	0.65	0.95	0.15	0.65	0.95	0.15	0.65	0.95
Right Triangular	0.2–0.9	0.1–0.9	0.1–0.9	0.2–0.9	0.1–0.9	0.1–0.9	0.2–0.9	0.1–0.9	0.1–0.9
Exponential	0.2–0.9	0.1–0.9	0.1–0.9	0.2–0.9	0.1–0.9	0.1–0.9	0.1–0.9	0.1–0.9	0.1–0.9
Normal	0.2–0.9	0.2–0.9	0.2–0.9	0.2–0.9	0.2–0.9	0.2–0.9	0.2–0.9	0.2–0.9	0.2–0.9
Chi-sq, $\nu = 6$	0.6–0.9	0.5–0.9	0.1–0.9	0.5–0.9	0.5–0.9	0.1–0.9	0.5–0.9	0.4–0.9	0.1–0.9
Gamma (2, 1)	0.2–0.9	0.1–0.9	0.1–0.9	0.2–0.9	0.1–0.9	0.1–0.9	0.1–0.9	0.1–0.9	0.1–0.9
Log Normal	0.2–0.9	0.1–0.9	0.1–0.9	0.1–0.9	0.1–0.9	0.1–0.9	0.1–0.9	0.1–0.9	0.1–0.9
Beta (3, 2)	0.5–0.9	0.3–0.9	0.1–0.9	0.4–0.9	0.3–0.9	0.1–0.9	0.4–0.9	0.3–0.9	0.1–0.9

or mderate, then a high value of correlation coefficient ρ is required for the proposed estimators to efiicient in case of Chi-Square ($\nu = 6$) and Beta (3, 2) distribution. If the value of ρ^* is high then the proposed estimator remains always more efficient. Interestingly, the choice of unrelated variable in the randomization device could be decided based on its correlation with the auxiliary variable used in the selection stage. Table 2 indicates that the value ρ could be any value in the range [0.1, 0.9], the the proposed estimators performs better for $\rho^* = 0.95$ in case of all the seven distributions considered in the simulation study.

Acknowledgements

The authors are thankful to the Editor, Associate Editor and two referees for their comments on the original version of the manuscript.

Appendix A:

Proof: (Proof of Theorem 1.) Let E_1 and E_2 denote the expected values with respect to sampling design and over randomization device respectively and let V_1 and V_2 be the corresponding variances, then

$$V\left(\widehat{Y}_1\right) = E_1 V_2\left(\widehat{Y}_1\right) + V_1 E_2\left(\widehat{Y}_1\right)$$
$$= E_1 V_2\left[\frac{1}{T}\left\{\frac{1}{n}\sum_{i=1}^{n}\frac{r_i}{p_i} - (1-T)U\right\}\right] + V_1 E_2\left[\frac{1}{T}\left\{\frac{1}{n}\sum_{i=1}^{n}\frac{r_i}{p_i} - (1-T)U\right\}\right].$$

On using (4.3) and substituting $E(r_i) = y_iT + (1-T)u_i$, we have

$$V\left(\widehat{Y}_1\right) = E_1\left[\frac{1}{T^2}\left\{\frac{1}{n^2}\sum_{i=1}^{n}\frac{T(1-T)(y_i-u_i)^2}{p_i^2}\right\}\right] + V_1\left[\frac{1}{T}\left\{\frac{1}{n}\sum_{i=1}^{n}\frac{y_iT+(1-T)u_i}{p_i} - (1-T)U\right\}\right]$$
$$= \frac{1-T}{nT}\sum_{i=1}^{N}\frac{(Y_i-U_i)^2}{p_i} + \frac{1}{nT^2}\sum_{i=1}^{N}p_i\left\{T\left(\frac{Y_i}{p_i}-Y\right)+(1-T)\left(\frac{U_i}{p_i}-U\right)\right\}^2.$$

□

Full derivation of the equation 5.3:

Let E_1 and E_2 denote the expected values with respect to sampling design and over randomization device respectively and let V_1 and V_2 be the corresponding variances, then

$$V\left(\widehat{Y}_2\right)_h = E_1V_2\left(\widehat{Y}_2\right) + V_1E_2\left(\widehat{Y}_2\right)$$
$$= E_1V_2\left[\frac{1}{T}\left\{\sum_{i=1}^{n}\frac{r_i}{p_i^*} - (1-T)U\right\}\right] + V_1E_2\left[\frac{1}{T}\left\{\frac{1}{n}\sum_{i=1}^{n}\frac{r_i}{p_i^*} - (1-T)U\right\}\right]$$
$$= E_1\left[\frac{1}{T^2}\left\{\frac{1}{n^2}\sum_{i=1}^{n}\frac{V_2(r_i)}{p_{hi}^{*2}}\right\}\right] + V_1\left[\frac{1}{T}\left\{\frac{1}{n}\sum_{i=1}^{n}\frac{E_2(r_i)}{p_{ih}^*} - (1-T)U\right\}\right].$$

On using (4.3) and substituting $E(r_i) = y_iT + (1-T)u_i$, we have

$$V\left(\widehat{Y}_2\right)_h = E_1\left[\frac{1}{T^2}\left\{\frac{1}{n^2}\sum_{i=1}^{n}\frac{T(1-T)(y_i-u_i)^2}{p_{ih}^{*2}}\right\}\right] + V_1\left[\frac{1}{T}\left\{\frac{1}{n}\sum_{i=1}^{n}\frac{y_iT+(1-T)u_i}{p_{ih}^*} - (1-T)U\right\}\right]$$
$$= E_1\left[\frac{1}{T^2}\left\{\frac{1}{n^2}\sum_{i=1}^{n}\frac{T(1-T)(y_i-u_i)^2}{p_{ih}^{*2}}\right\}\right] + \frac{1}{n^2T^2}V_1\left\{\sum_{i=1}^{n}\frac{y_iT+(1-T)u_i}{p_{ih}^*}\right\}$$
$$= \frac{1-T}{nT}\sum_{i=1}^{N}\frac{(Y_i-U_i)^2p_i}{p_{ih}^{*2}} + \frac{1}{nT^2}\sum_{i=1}^{N}p_i\left\{\frac{TY_i+(1-T)U_i}{p_{ih}^*} - \sum_{i=1}^{N}\frac{TY_i+(1-T)U_i}{p_{ih}^*}p_i\right\}^2.$$
$$= \frac{1-T}{nT}\sum_{i=1}^{N}\frac{(Y_i-U_i)^2p_i}{p_{ih}^{*2}} + \frac{1}{nT^2}\left[\sum_{i=1}^{N}\frac{(TY_i+(1-T)U_i)^2p_i}{p_{ih}^{*2}} - \left\{\sum_{i=1}^{N}\frac{(TY_i+(1-T)U_i)p_i}{p_{ih}^*}\right\}^2\right],$$

which proves the equation.

Proof: (Proof of Theorem 2.) We know that

$$E_m\left\{\text{MSE}\left(\widehat{Y}_2\right)_h\right\} = E_m\left\{\text{V}\left(\widehat{Y}_2\right)\right\} + E_m\left\{\text{B}\left(\widehat{Y}_2\right)_h\right\}^2.$$

Thus we have

$$E_m\left\{\mathrm{MSE}\left(\widehat{Y}_2\right)_h\right\} = \frac{(1-T)\,E_m(B_1)}{nT} + \frac{E_m(B_2) - E_m(B_3)}{nT^2} + E_m\left\{B\left(\widehat{Y}_2\right)_h\right\}^2, \qquad \text{(A.1)}$$

where

$$B_1 = \sum_{i=1}^{N} \frac{(Y_i - U_i)^2}{p_{ih}^{*2}} p_i, \qquad B_2 = \sum_{i=1}^{N} \frac{\{TY_i + (1-T)\,U_i\}^2}{p_{ih}^{*2}} p_i$$

and

$$B_3 = \left\{\sum_{i=1}^{N} \frac{TY_i + (1-T)\,U_i}{p_{ih}^{*}} p_i\right\}^2.$$

Under the superpopulation model, we have

$$B_1 = \sum_{i=1}^{N} \frac{\left(\beta p_i + e_i - \beta^* p_i - e_i^*\right)^2}{p_{ih}^{*2}} p_i = \sum_{i=1}^{N} \frac{(\beta-\beta^*)^2\, p_i^2 + \left(e_i - e_i^*\right)^2 + 2\,(\beta-\beta^*)\, p_i \left(e_i - e_i^*\right)}{p_{ih}^{*2}} p_i.$$

Thus

$$E_m(B_1) = (\beta-\beta^*)^2 \sum_{i=1}^{N} \frac{p_i^3}{p_{ih}^{*2}} + 2a \sum_{i=1}^{N} \frac{p_i^{g+1}}{p_{ih}^{*2}}.$$

Now

$$\begin{aligned}
B_2 &= \left\{\sum_{i=1}^{N} \frac{T^2 Y_i^2 + (1-T)^2\, U_i^2 + 2T(1-T) Y_i U_i}{p_{ih}^{*2}} p_i\right\} \\
&= \sum_{i=1}^{N} \frac{T^2\,(\beta p_i + e_i)^2 + (1-T)^2 \left(\beta^* p_i + e_i^*\right)^2 + 2T\,(1-T)\,(\beta p_i + e_i)\left(\beta^* p_i + e_i^*\right)}{p_{ih}^{*2}} p_i \\
&= \sum_{i=1}^{N} \frac{T^2\left(\beta^2 p_i^2 + e_i^2 + 2\beta p_i e_i\right) + (1-T)^2\left(\beta^{*2} p_i^2 + e_i^{*2} + 2\beta^* p_i e_i^*\right)}{p_{ih}^{*2}} p_i \\
&\quad + 2\sum_{i=1}^{N} \frac{T\,(1-T)\left(\beta\beta^* p_i^2 + \beta p_i e_i^* + \beta^* p_i e_i + e_i e_i^*\right)}{p_{ih}^{*2}} p_i.
\end{aligned}$$

Thus, we have

$$E_m(B_2) = \left\{T^2 + (1-T)^2\right\} a \sum_{i=1}^{N} \frac{p_i^{g+1}}{p_{ih}^{*2}} + \{\beta T + (1-T)\,\beta^*\}^2 \sum_{i=1}^{N} \frac{p_i^3}{p_{ih}^{*2}}.$$

Now

$$B_3 = \left\{\sum_{i=1}^{N} \frac{TY_i + (1-T)\,U_i}{p_{ih}^{*}} p_i\right\}^2 = \left\{T \sum_{i=1}^{N} \frac{Y_i p_i}{p_{ih}^{*}} + (1-T) \sum_{i=1}^{N} \frac{U_i p_i}{p_{ih}^{*}}\right\}^2.$$

Using the superpopulation models, we have Using the superpopulation models, we have

$$B_3 = T^2\left\{\sum_{i=1}^{N}\frac{(\beta p_i + e_i)\,p_i}{p^*_{ih}}\right\}^2 + (1-T)^2\left\{\sum_{i=1}^{N}\frac{\left(\beta^* p_i + e^*_i\right)p_i}{p^*_{ih}}\right\}^2$$
$$+ 2T(1-T)\left\{\sum_{i=1}^{N}\frac{\left(\beta^* p_i + e^*_i\right)p_i}{p^*_{ih}}\right\}\left\{\sum_{i=1}^{N}\frac{(\beta p_i + e_i)\,p_i}{p^*_{ih}}\right\}$$
$$= T^2\left(\beta\sum_{i=1}^{N}\frac{p_i^2}{p^*_{ih}} + \sum_{i=1}^{N}\frac{e_i p_i}{p^*_{ih}}\right)^2 + (1-T)^2\left(\beta^*\sum_{i=1}^{N}\frac{p_i^2}{p^*_{ih}} + \sum_{i=1}^{N}\frac{e^*_i p_i}{p^*_{ih}}\right)^2$$
$$+ 2T(1-T)\left(\beta\sum_{i=1}^{N}\frac{p_i^2}{p^*_{ih}} + \sum_{i=1}^{N}\frac{e_i p_i}{p^*_{ih}}\right)\left(\beta^*\sum_{i=1}^{N}\frac{p_i^2}{p^*_{ih}} + \sum_{i=1}^{N}\frac{e^*_i p_i}{p^*_{ih}}\right)$$
$$= T^2\left\{\beta^2\left(\sum_{i=1}^{N}\frac{p_i^2}{p^*_{ih}}\right)^2 + \sum_{i=1}^{N}\frac{e_i^2 p_i^2}{p^{*2}_{ih}} + \sum_{j\neq i=1}^{N}\frac{p_i p_j e_i e_j}{p^*_{ih}p^*_{jh}} + 2\beta\left(\sum_{i=1}^{N}\frac{p_i^2}{p^*_{ih}}\sum_{i=1}^{N}\frac{p_i e_i}{p^*_{ih}}\right)\right\}$$
$$+ (1-T)^2\left\{\beta^{*2}\left(\sum_{i=1}^{N}\frac{p_i^2}{p^*_{ih}}\right)^2 + \sum_{i=1}^{N}\frac{e_i^{*2} p_i^2}{p^{*2}_{ih}} + \sum_{j\neq i=1}^{N}\frac{p_i p_j e^*_i e^*_j}{p^*_{ih}p^*_{jh}} + 2\beta^*\left(\sum_{i=1}^{N}\frac{p_i^2}{p^*_{ih}}\sum_{i=1}^{N}\frac{p_i e^*_i}{p^*_{ih}}\right)\right\}$$
$$+ 2T(1-T)\left\{\beta\beta^*\left(\sum_{i=1}^{N}\frac{p_i^2}{p^*_{ih}}\right)^2 + \beta\left(\sum_{i=1}^{N}\frac{p_i^3 e^*_i}{p^{*2}_{ih}} + \sum_{j\neq i=1}^{N}\frac{p_i^2 e^*_j p_j}{p^*_{ih}p^*_{jh}}\right)\right.$$
$$\left. +\beta^*\left(\sum_{i=1}^{N}\frac{p_i^3 e_i}{p^{*2}_{ih}} + \sum_{j\neq i=1}^{N}\frac{p_i^2 e_j p_j}{p^*_{ih}p^*_{jh}}\right) + \left(\sum_{i=1}^{N}\frac{e_i e^*_i p_i^2}{p^{*2}_{ih}} + \sum_{j\neq i=1}^{N}\frac{e_i e^*_j p_i p_j}{p^*_{ih}p^*_{jh}}\right)\right\}.$$

Taking expected value, we have

$$E_m(B_3) = T^2\left\{\beta^2\left(\sum_{i=1}^{N}\frac{p_i^2}{p^*_{ih}}\right)^2 + a\sum_{i=1}^{N}\frac{p_i^{g+2}}{p^{*2}_{ih}}\right\} + (1-T)^2\left\{\beta^{*2}\left(\sum_{i=1}^{N}\frac{p_i^2}{p^*_{ih}}\right)^2 + a\sum_{i=1}^{N}\frac{p_i^{g+2}}{p^{*2}_{ih}}\right\}$$
$$+ 2T(1-T)\left\{\beta\beta^*\left(\sum_{i=1}^{N}\frac{p_i^2}{p^*_{ih}}\right)^2\right\}$$
$$= a\left\{T^2 + (1-T)^2\right\}\left(\sum_{i=1}^{N}\frac{p_i^{g+2}}{p^{*2}_{ih}}\right) + T^2\beta^2\left(\sum_{i=1}^{N}\frac{p_i^2}{p^*_{ih}}\right)^2 + (1-T)^2\beta^{*2}\left(\sum_{i=1}^{N}\frac{p_i^2}{p^*_{ih}}\right)^2$$
$$+ 2T(1-T)\beta\beta^*\left(\sum_{i=1}^{N}\frac{p_i^2}{p^*_{ih}}\right)^2$$
$$= a\left\{T^2 + (1-T)^2\right\}\left(\sum_{i=1}^{N}\frac{p_i^{g+2}}{p^{*2}_{ih}}\right) + \left[\left\{\beta T + (1-T)\beta^*\right\}\sum_{i=1}^{N}\frac{p_i^2}{p^*_{ih}}\right]^2.$$

Now we have

$$
\begin{aligned}
E_m\left\{B\left(\widehat{Y}_2\right)_h\right\}^2 &= E_m\left\{\sum_{i=1}^{N}\left(\frac{p_i}{p_{ih}^*}-1\right)\left(Y_i+\frac{1-T}{T}U_i\right)\right\}^2 \\
&= E_m\left\{\sum_{i=1}^{N}\left(\frac{p_i}{p_{ih}^*}-1\right)Y_i+\frac{1-T}{T}\sum_{i=1}^{N}\left(\frac{p_i}{p_{ih}^*}-1\right)U_i\right\}^2 \\
&= E_m\left[\left\{\sum_{i=1}^{N}\left(\frac{p_i}{p_{ih}^*}-1\right)Y_i\right\}^2+\frac{(1-T)^2}{T^2}\left\{\sum_{i=1}^{N}\left(\frac{p_i}{p_{ih}^*}-1\right)U_i\right\}^2\right. \\
&\quad\left.+\frac{2(1-T)}{T}\left\{\sum_{i=1}^{N}\left(\frac{p_i}{p_{ih}^*}-1\right)^2Y_iU_i+\sum_{j\neq i=1}^{N}\left(\frac{p_i}{p_{ih}^*}-1\right)\left(\frac{p_j}{p_{jh}^*}-1\right)Y_iU_j\right\}\right] \\
&= \beta^2\left(\sum_{i=1}^{N}\frac{p_i^2}{p_{ih}^*}-1\right)^2+a\sum_{i=1}^{N}\left(\frac{p_i^{g+2}}{p_{ih}^{*2}}+p_i^g-2\frac{p_i^{g+1}}{p_{ih}^*}\right) \\
&\quad+\frac{(1-T)^2}{T^2}\left\{\beta^{*2}\left(\sum_{i=1}^{N}\frac{p_i^2}{p_{ih}^*}-1\right)^2+a\sum_{i=1}^{N}\left(\frac{p_i^{g+2}}{p_{ih}^{*2}}+p_i^g-2\frac{p_i^{g+1}}{p_{ih}^*}\right)\right\} \\
&\quad+2\frac{1-T}{T}\left(\sum_{i=1}^{N}\frac{p_i^2}{p_{ih}^*}-1\right)^2\beta\beta^* \\
&= \left\{\beta\left(\sum_{i=1}^{N}\frac{p_i^2}{p_{ih}^*}-1\right)+\frac{1-T}{T}\beta^*\left(\sum_{i=1}^{N}\frac{p_i^2}{p_{ih}^*}-1\right)\right\}^2 \\
&\quad+a\left\{1+\frac{(1-T)^2}{T^2}\right\}\sum_{i=1}^{N}\left(\frac{p_i^{g+2}}{p_{ih}^{*2}}+p_i^g-2\frac{p_i^{g+1}}{p_{ih}^*}\right).
\end{aligned}
$$

On substituting $E(B_1)$, $E(B_2)$, $E(B_3)$ and $E_m(B(\widehat{Y}_2)_h)^2$ in $(A.1)$ and on re-arranging, we can get the following one:

$$
E_m\left[\mathrm{MSE}\left(\widehat{Y}_2\right)_h\right]=\frac{1}{n}A_1+A_2,
$$

where

$$
\begin{aligned}
A_1 &= \left[\left(\frac{1-T}{T}\right)\left\{(\beta-\beta^*)^2\sum_{i=1}^{N}\frac{p_i^3}{p_{ih}^{*2}}+2a\sum_{i=1}^{N}\frac{p_i^{g+1}}{p_{ih}^{*2}}\right\}\right. \\
&\quad+\frac{1}{T^2}\left\{a\left(T^2+(1-T)^2\right)\left(\sum_{i=1}^{N}\frac{p_i^{g+1}}{p_{ih}^{*2}}-\sum_{i=1}^{N}\frac{p_i^{g+2}}{p_{ih}^{*2}}\right)\right. \\
&\quad\left.\left.+(\beta T+(1-T)\beta^*)^2\left(\sum_{i=1}^{N}\frac{p_i^3}{p_{ih}^{*2}}-\left(\sum_{i=1}^{N}\frac{p_i^2}{p_{ih}^*}\right)^2\right)\right\}\right]
\end{aligned}
$$

and

$$A_2 = \left\{\left(\sum_{i=1}^{N} \frac{p_i^2}{p_{ih}^*} - 1\right)\left(\beta + \frac{(1-T)}{T}\beta^*\right)\right\}^2 + a\left\{1 + \frac{(1-T)^2}{T^2}\right\}\sum_{i=1}^{N}\left(\frac{p_i^{g+2}}{p_{ih}^{*2}} + p_i^g - 2\frac{p_i^{g+1}}{p_{ih}^*}\right).$$

□

Proof: (Proof of Theorem 3.) Let C_1 be the cost per unit of collecting information of each individual. The cost C_0 of observing the sample of size n is given by

$$C_0 = nC_1. \tag{A.2}$$

To minimize $E_m[\text{MSE}(\widehat{Y}_2)_h]$ subject to condition (A.2), consider the function

$$L = \frac{1}{n}A_1 + A_2 + \lambda(nC_1 - C_0). \tag{A.3}$$

Differentiating (A.3) partially with respect to n and we get $n = C_0/C_1$ and hence (5.4) becomes (5.7), which proves the theorem.

□

Appendix B:

Relative efficiencies(RE_h), h = 1, 2, 3, 4, 5 of the proposed estimators for PPSWR sampling scheme using Unrelated Question Model under various distributions.

		$\rho^* = 0.15$					$\rho^* = 0.65$					$\rho^* = 0.95$				
g	ρ	RE_1	RE_2	RE_3	RE_4	RE_5	RE_1	RE_2	RE_3	RE_4	RE_5	RE_1	RE_2	RE_3	RE_4	RE_5
		Right Triangular														
0	.1	75.76	75.78	72.30	68.25	93.72	106.4	106.5	99.23	93.15	180.2	124.8	124.9	115.5	108.6	258.7
	.2	100.0	100.1	90.32	79.20	122.2	142.9	143.0	124.2	107.9	237.2	165.7	165.9	143.6	126.8	334.9
	.3	140.1	140.2	119.4	96.95	165.0	201.0	201.3	163.1	130.1	308.8	224.2	224.5	183.9	152.1	416.0
	.4	200.2	200.4	161.9	122.6	221.9	293.5	294.0	222.8	163.2	401.0	310.9	311.5	243.6	189.0	514.2
	.5	285.0	285.4	220.9	159.3	292.6	439.2	440.1	313.8	213.2	517.8	446.2	447.0	336.4	245.6	643.9
	.6	398.3	398.8	300.8	213.1	378.2	660.6	661.9	449.4	291.4	662.5	672.1	673.3	489.8	339.2	831.4
	.7	540.4	541.0	407.6	296.1	482.5	964.6	966.2	643.6	420.0	838.8	1089	1091	766.4	513.4	1135
	.8	706.6	707.2	553.0	434.7	617.0	1297	1298	904.3	645.1	1053	1977	1979	1338	905.4	1717
	.9	897.9	898.3	775.1	691.5	818.0	1515	1516	1231	1051	1312	3900	3904	2726	2083	3032
1	.1	76.51	76.54	73.08	70.46	113.4	102.8	102.8	96.15	91.33	187.8	123.3	123.4	114.2	107.8	259.4
	.2	100.6	100.7	91.85	84.63	150.9	138.2	138.3	121.4	108.6	257.4	163.8	164.0	142.3	126.5	340.6
	.3	140.1	140.2	122.4	107.1	205.7	194.8	195.0	161.3	135.2	350.6	221.7	222.1	182.9	153.4	430.6
	.4	200.9	201.1	169.1	140.6	281.2	286.4	286.9	225.2	176.7	481.1	308.3	308.8	244.1	193.4	545.1
	.5	290.8	291.2	238.9	190.6	379.9	436.8	437.7	329.7	243.1	664.7	445.2	446.1	342.0	257.2	706.5
	.6	418.3	418.8	342.0	266.6	503.8	683.0	684.4	504.3	355.1	919.3	680.9	682.2	513.2	368.9	961.5
	.7	589.1	589.7	491.3	386.5	654.3	1065	1067	796.6	555.3	1255	1145	1147	857.0	595.6	1433
	.8	799.4	799.9	698.9	580.5	832.3	1552	1553	1245	924.6	1634	2295	2298	1732	1189	2546
	.9	1029	1030	963.8	883.0	1035	1867	1868	1698	1485	1868	5883	5887	4752	3511	5856
2	.1	77.93	77.97	74.07	71.38	128.3	101.9	102.0	95.21	90.50	197.1	122.8	122.9	113.7	107.3	261.9
	.2	103.2	103.3	93.63	86.43	173.7	137.9	138.0	120.9	108.4	276.9	163.3	163.5	141.9	126.3	346.8
	.3	144.1	144.2	125.2	110.1	238.2	195.5	195.8	161.7	136.2	385.9	221.6	222.0	182.7	153.5	442.4
	.4	207.5	207.8	174.2	145.9	326.9	290.0	290.5	227.8	179.8	543.1	309.3	309.8	244.8	194.4	566.1
	.5	303.0	303.4	248.8	200.0	442.8	448.3	449.2	338.3	251.0	772.5	449.2	450.0	345.0	260.0	743.8
	.6	440.8	441.4	361.6	283.7	586.4	716.3	717.8	529.2	373.8	1101	693.8	695.1	523.2	376.6	1033
	.7	627.9	628.6	528.0	417.0	755.0	1151	1153	864.2	600.1	1538	1190	1192	892.2	618.9	1590
	.8	855.9	856.5	758.6	631.3	939.5	1712	1723	1399	1029	1988	2497	2501	1895	1284	3018
	.9	1088	1088	1031	946.0	1119	2038	2039	1885	1645	2117	7194	7199	5890	4182	7886

		ρ^* = 0.15					ρ^* = 0.65					ρ^* = 0.95				
g	ρ	RE_1	RE_2	RE_3	RE_4	RE_5	RE_1	RE_2	RE_3	RE_4	RE_5	RE_1	RE_2	RE_3	RE_4	RE_5
									Exponential							
0	.1	86.55	86.70	78.87	73.71	107.1	127.3	127.7	108.3	97.86	249.4	150.5	151.1	124.4	110.9	430.8
	.2	124.8	125.2	104.1	89.55	145.1	191.9	192.9	144.5	117.1	328.3	230.8	232.3	167.2	133.3	576.5
	.3	187.9	188.6	145.6	114.7	204.0	292.4	294.3	201.1	146.2	424.3	349.2	351.7	231.1	165.3	721.1
	.4	282.6	283.7	208.7	151.8	286.0	446.6	449.6	289.4	190.4	547.8	526.6	530.6	329.3	213.0	888.2
	.5	416.4	418.1	302.7	207.6	396.4	675.2	679.4	428.3	260.2	711.5	797.0	802.8	487.1	288.8	1103
	.6	597.5	599.5	442.7	295.7	546.0	997.2	1002	648.8	376.9	936.6	1219	1226	755.5	420.1	1413
	.7	837.0	839.1	653.6	445.3	756.5	1426	1431	1001	589.0	1263	1910	1919	1254	679.4	1923
	.8	1163	1165	982.6	725.4	1073	1980	1983	1566	1019	1765	3195	3203	2325	1315	2951
	.9	1661	1663	1538	1316	1605	2699	2701	2447	1971	2560	6132	6138	5217	3562	5640
1	.1	94.12	94.36	83.41	77.36	165.8	124.5	125.0	105.6	95.6	290.5	148.8	149.4	123.0	109.8	443.6
	.2	136.7	137.3	111.0	94.78	223.4	190.2	191.2	142.9	116.3	406.7	228.9	230.3	165.9	132.5	611.7
	.3	205.1	206.1	155.5	121.7	305.2	294.5	296.5	201.9	147.4	550.3	348.1	350.7	230.4	165.1	788.0
	.4	311.4	312.9	225.4	162.3	417.9	462.1	465.3	296.9	195.1	739.3	530.6	534.7	331.1	214.1	1001
	.5	470.0	472.2	334.4	224.8	569.1	729.5	734.5	453.8	271.9	994.3	819.1	825.3	496.8	293.0	1288
	.6	696.4	699.2	504.1	325.9	769.9	1145	1152	720.1	404.0	1346	1297	1306	790.4	432.3	1719
	.7	1005	1008	768.4	501.3	1039	1754	1761	1180	653.5	1838	2158	2169	1373	716.9	2468
	.8	1409	1412	1176	833.2	1408	2552	2558	1954	1182	2521	3989	4002	2775	1458	4086
	.9	1939	1940	1791	1503	1928	3358	3361	3020	2340	3312	8786	8796	7192	4415	8549
2	.1	102.6	103.0	87.78	80.25	239.7	125.3	125.8	104.8	94.38	345.4	148.0	148.6	122.0	108.9	462.0
	.2	152.9	153.7	118.4	98.95	333.7	195.2	196.4	143.6	116.0	511.5	229.1	230.6	165.3	131.8	653.5
	.3	231.1	232.5	167.0	127.4	456.0	307.3	309.6	205.4	148.4	724.8	351.1	353.8	230.7	164.9	862.8
	.4	354.0	356.1	244.5	170.9	620.0	493.3	497.3	307.0	198.7	1018	541.2	545.6	334.0	214.9	1126
	.5	543.4	546.6	369.1	239.1	834.6	807.5	814.1	481.4	280.8	1430	850.5	857.3	507.3	296.2	1495
	.6	824.5	828.7	570.7	351.9	1107	1337	1347	794.2	425.8	2009	1389	1399	824.3	441.9	2076
	.7	1214	1219	894.6	552.3	1443	2186	2199	1376	709.4	2795	2439	2453	1489	747.1	3158
	.8	1702	1706	1391	936.4	1840	3324	3335	2415	1337	3711	4996	5016	3267	1583	5760
	.9	1661	1663	1538	1316	1605	2699	2701	2447	1971	2560	6132	6138	5217	3562	5640

		ρ^* = 0.15					ρ^* = 0.65					ρ^* = 0.95				
g	ρ	RE_1	RE_2	RE_3	RE_4	RE_5	RE_1	RE_2	RE_3	RE_4	RE_5	RE_1	RE_2	RE_3	RE_4	RE_5
									Normal							
0	.1	80.39	80.47	75.32	71.17	98.84	114.7	115.0	103.0	95.27	206.0	135.2	135.5	119.2	109.5	321.4
	.2	109.7	109.9	95.95	84.60	129.6	161.9	162.4	132.5	112.3	269.2	191.3	192.0	153.6	129.7	421.2
	.3	158.1	158.5	129.7	106.3	176.6	236.0	237.0	178.4	138.3	346.8	271.9	273.1	203.6	158.2	523.3
	.4	230.7	231.4	180.2	138.4	240.8	351.6	353.2	249.6	177.9	446.3	391.3	393.1	278.7	200.5	643.9
	.5	332.9	333.9	253.2	185.7	324.7	527.8	530.2	360.2	239.7	576.2	574.7	577.4	397.0	266.8	801.5
	.6	469.2	470.5	357.9	257.9	434.0	783.6	786.8	532.1	340.8	749.6	870.3	874.1	595.7	380.3	1031
	.7	643.1	644.5	508.2	374.4	580.7	1124	1127	797.4	517.7	989.6	1384	1389	963.6	601.3	1413
	.8	862.4	863.5	728.8	576.9	789.5	1526	1529	1196	850.6	1334	2401	2407	1766	1133	2196
	.9	1162	1162	1073	957.3	1115	1934	1935	1741	1482	1813	4736	4741	3948	2929	4240
1	.1	84.08	84.19	77.92	73.87	134.3	111.7	112.0	100.4	93.26	227.1	133.7	134.0	118.0	108.6	326.6
	.2	115.3	115.5	100.2	89.33	179.2	159.0	159.5	130.6	111.9	313.9	189.5	190.2	152.5	129.1	437.6
	.3	165.7	166.1	136.2	113.5	243.6	233.8	234.8	178.4	140.4	425.3	270.2	271.4	202.9	158.4	557.1
	.4	243.5	244.3	192.0	149.9	331.9	354.5	356.1	254.9	184.4	576.0	391.4	393.3	280.0	202.3	705.4
	.5	358.9	360.1	276.7	204.9	448.2	550.0	552.6	380.1	255.0	782.4	582.5	585.3	404.3	272.5	909.4
	.6	522.5	524.0	404.4	291.1	597.4	861.7	865.4	590.1	375.0	1066	905.3	909.4	622.0	395.8	1224
	.7	741.1	742.7	594.5	432.9	787.3	1328	1332	944.6	595.1	1450	1515	1521	1055	646.4	1788
	.8	1014	1016	870	678.4	1029	1915	1919	1508	1027	1932	2917	2924	2126	1299	3068
	.9	1334	1335	1245	1104	1333	2374	2376	2158	1802	2355	6885	6893	5655	3865	6747
2	.1	88.60	88.76	80.63	75.9	170.6	111.6	111.8	99.47	92.26	252.1	133.0	133.3	117.2	107.8	334.1
	.2	123.4	123.7	104.8	92.51	232.9	160.8	161.3	130.8	111.8	361.9	189.3	190.0	117.2	107.8	334.1
	.3	178.1	178.7	143.1	118.2	317.3	239.1	240.2	180.4	141.6	507.2	271.2	272.5	203.0	158.4	587.4
	.4	263.0	264.0	203.3	157.1	431.7	367.7	369.5	261.0	187.9	711.8	395.5	397.5	281.7	203.3	757.0
	.5	391.8	393.3	296.9	217.0	580.9	583.5	586.6	397.0	263.6	1004	594.8	597.8	410.4	275.6	997.5
	.6	579.2	581.1	442.3	312.5	767.1	947.1	951.7	635.5	395.5	1417	1630	1637	1120	673.1	2115
	.7	833.8	835.8	663.5	472.3	989.3	1529	1534	1062	645.3	1972	1630	1637	1120	673.1	2115
	.8	1144	1145	980.3	749.2	1239	2284	2289	1771	1154	2570	3378	3388	2407	1408	3940
	.9	1462	1462	1368	1204	1493	2733	2735	2487	2036	2812	9293	9305	7402	4646	9918

		$\rho^* = 0.15$					$\rho^* = 0.65$					$\rho^* = 0.95$				
g	ρ	RE_1	RE_2	RE_3	RE_4	RE_5	RE_1	RE_2	RE_3	RE_4	RE_5	RE_1	RE_2	RE_3	RE_4	RE_5
		Chi Square														
0	.1	63.36	63.37	63.03	62.66	65.58	91.89	91.91	91.38	90.83	99.67	100.8	100.8	100.3	99.79	110.0
	.2	66.06	66.08	65.28	64.40	68.46	92.64	92.67	91.51	90.27	100.6	103.8	103.9	102.9	101.8	112.8
	.3	75.05	75.08	73.71	72.18	77.84	94.28	94.33	92.36	90.18	101.9	107.1	107.2	105.7	104.2	115.1
	.4	87.37	87.41	85.52	83.40	90.35	98.58	98.65	95.58	92.08	105.8	110.8	110.9	109.1	107.2	117.5
	.5	99.55	99.60	97.48	95.04	102.3	109.9	110.0	105.3	99.85	116.7	115.4	115.5	113.5	111.2	120.8
	.6	109.7	109.8	107.8	105.4	111.9	134.0	134.1	127.8	119.9	140.1	122.2	122.3	120.1	117.4	126.3
	.7	117.5	117.5	115.8	113.9	118.9	158.9	159.0	153.5	146.0	162.6	135.6	135.7	133.3	130.2	138.7
	.8	123.0	123.0	121.8	120.5	123.7	159.1	159.2	1567	153.2	160.5	182.5	182.6	179.4	174.8	185.3
	.9	126.5	126.5	125.9	125.2	126.7	145.9	145.9	145.1	144.2	146.2	642.3	642.4	629.7	604.7	651.2
1	.1	63.68	63.69	63.38	63.05	68.05	91.50	91.71	91.20	90.67	100.3	100.8	100.8	100.3	99.76	110.1
	.2	66.84	66.86	66.11	65.33	71.78	92.66	92.69	91.55	90.35	101.9	103.8	103.9	102.9	101.8	113.0
	.3	76.08	76.11	74.82	73.42	81.48	94.65	94.70	92.79	90.70	104.0	107.1	107.2	105.8	104.3	115.3
	.4	88.54	88.59	86.81	84.82	93.85	99.56	99.64	96.68	93.3	109.3	110.9	111.0	109.2	107.3	117.8
	.5	100.8	100.9	98.84	96.51	105.3	111.9	112.1	107.6	102.2	122.4	115.6	115.7	113.7	111.4	121.3
	.6	110.9	111.0	109.1	106.8	114.3	137.8	137.9	131.8	124.0	148.7	122.5	122.6	120.4	117.8	127.0
	.7	118.5	118.6	117.0	115.0	120.6	163.4	163.5	158.3	150.8	170.7	136.3	136.4	134.1	131.0	139.9
	.8	123.7	123.8	122.6	121.3	124.7	161.8	161.9	159.5	156.1	164.3	184.6	184.7	181.7	177.1	188.5
	.9	126.9	126.9	126.3	125.6	127.2	146.8	146.8	146.0	145.1	147.3	676.0	676.2	663.9	637.5	639.7
2	.1	64.00	64.00	63.70	63.40	70.13	91.56	91.58	91.07	90.56	100.9	100.8	100.8	100.2	99.73	110.2
	.2	67.52	67.54	66.82	66.08	74.59	92.70	92.73	91.61	90.45	102.9	103.8	103.9	102.9	101.8	113.1
	.3	76.98	77.01	75.76	74.42	84.54	95.00	95.04	93.16	91.12	105.8	107.2	107.2	105.8	104.3	115.5
	.4	89.57	89.61	87.88	85.94	96.77	100.4	100.5	97.58	94.26	112.2	111.0	111.0	109.3	107.4	118.1
	.5	101.9	101.9	99.96	97.68	107.8	113.7	113.8	109.4	104.1	127.2	115.8	115.8	113.8	111.6	121.6
	.6	112.0	112.0	110.1	107.9	116.2	141.0	141.1	135.1	127.2	155.8	122.8	122.9	120.7	118.1	127.6
	.7	119.4	119.4	117.8	116.0	121.9	167.1	167.2	162.1	154.6	177.3	136.9	136.9	134.6	131.6	140.9
	.8	124.3	124.4	123.2	121.9	125.5	164.0	164.1	161.8	158.4	167.4	186.4	186.4	183.5	178.9	190.9
	.9	127.2	127.2	126.6	126.0	127.5	147.5	147.5	146.8	145.8	148.1	703.7	703.8	691.7	663.9	728.7

		$\rho^* = 0.15$					$\rho^* = 0.65$					$\rho^* = 0.95$				
g	ρ	RE_1	RE_2	RE_3	RE_4	RE_5	RE_1	RE_2	RE_3	RE_4	RE_5	RE_1	RE_2	RE_3	RE_4	RE_5
		Gamma														
0	.1	82.76	82.93	77.84	73.73	99.20	113.9	114.4	102.9	95.46	202.4	133.6	134.2	118.5	109.1	307.0
	.2	110.4	110.8	97.49	86.24	134.2	158.4	159.4	131.3	111.9	263.4	186.6	188.0	151.6	128.7	398.2
	.3	156.6	156.4	129.6	106.6	179.5	227.6	229.4	175.4	136.9	338.8	261.5	263.8	199.3	156.3	492.4
	.4	223.5	224.8	177.7	136.7	241.1	335.2	338.2	243.6	174.8	435.9	371.1	374.6	270.5	197.0	604.5
	.5	319.3	321.2	246.9	181.0	320.7	500.0	504.5	349.2	233.6	561.8	538.2	543.2	381.9	260.4	751.8
	.6	447.7	450.1	344.7	248.1	422.1	742.5	748.4	511.7	329.0	726.4	808.5	815.3	567.7	368.2	966.1
	.7	611.2	613.7	482.2	355.2	554.3	1070	1076	757.9	493.2	946.4	1286	1295	909.4	576.2	1322
	.8	812.3	814.5	678.8	538.0	737.2	1449	1455	1116	795.8	1249	2257	2268	1647	1071	2043
	.9	1071	1073	979.2	875.4	1018	1790	1792	1583	1352	1650	4466	4477	3608	2700	3876
1	.1	88.82	88.47	81.60	77.30	99.70	112.4	112.9	101.2	94.2	226.2	132.5	133.1	117.5	108.3	313.1
	.2	119.0	119.5	103.6	92.41	186.6	158.1	159.1	130.8	112.6	308.4	185.6	186.9	150.8	128.4	413.9
	.3	167.2	168.1	138.5	115.9	247.7	229.2	231.1	177.2	140.5	414.0	261.1	263.4	199.2	156.9	522.5
	.4	240.4	241.9	192.1	151.2	331.0	342.4	345.5	251.1	183.6	557.7	372.7	376.3	272.5	199.6	658.0
	.5	347.7	349.9	273.0	204.4	440.2	524.3	529.2	371.8	252.8	755.9	546.7	551.9	390.1	267.6	846.0
	.6	498.3	501.1	393.7	287.3	579.2	813.6	820.6	573.2	370.3	1029	839.3	846.5	595.0	386.8	1138
	.7	698.5	701.4	570.7	422.0	753.0	1248	1256	910.3	584.0	1394	1396	1406	1002	629.0	1669
	.8	946.2	948.6	821.3	649.9	968.4	1795	1802	1435	995.6	1837	2704	2718	2015	1261	2888
	.9	1229	1230	1151	1030	1230	2194	2197	2003	1693	2183	6530	6545	5387	3742	6436
2	.1	94.97	95.34	85.91	80.53	99.80	113.8	114.3	101.4	93.96	258.3	132.2	132.8	117.0	107.8	322.7
	.2	131.0	131.8	110.6	97.17	252.2	162.6	163.8	132.5	113.4	364.6	186.2	187.6	150.7	128.1	432.6
	.3	185.3	186.6	148.7	122.4	332.7	238.8	241.0	181.3	142.8	502.8	263.5	265.9	200.0	157.2	553.4
	.4	267.2	269.2	207.4	160.4	439.0	361.8	365.5	260.1	188.5	696.0	378.9	382.6	275.2	200.9	707.3
	.5	388.3	391.3	297.6	218.7	575.2	565.4	571.4	392.0	263.0	971.0	561.5	567.0	397.4	271.2	926.1
	.6	560.7	564.4	435.1	310.9	742.5	904.2	913.1	621.3	392.4	1358	876.2	884.3	615.1	396.2	1278
	.7	790.0	793.9	639.8	462.7	938.8	1440	1451	1026	635.5	1869	1501	1513	1065	656.0	1950
	.8	1063	1066	924.2	718.7	1155	2122	2132	1680	1119	2402	3102	3120	2273	1367	3642
	.9	1337	1338	1259	1120	1368	2496	2500	2291	1906	2575	8617	8640	6980	4492	9258

References

Amahia, G. N., Chaubey, Y. P. and Rao, T. J. (1989). Efficiency of a new estimator in PPS sampling for multiple characteristics, *Journal of Statistical Planning and Inference*, **21**, 72–84.

Arnab, R. (2001). Estimation of a finite population total in varying probability sampling for multi-character surveys, *Metrika*, **54**, 159–177.

Bansal, M. L. and Singh, R. (1985). An alternative estimator for multiple characteristics in PPS sampling, *Journal of Statistical Planning and Inference*, **11**, 313–20.

Bansal, M. L., Singh, S. and Singh, R. (1994). Multi-character survey using randomized response technique, *Communications in Statistics - Theory and Methods*, **23**, 1705–1715.

Greenberg, B. G., Kuebler, R. R., Abernathy, J. R. and Horvitz, D. G. (1971). Application of the randomized response technique in obtaining quantitative data, *Journal of the American Statistical Association*, **66**, 243–250.

Grewal, I. S., Bansal, M. L. and Singh, S. (1997). An alternative estimator for multiple characteristics using randomized response technique in pps sampling, *The Aligarh Journal of Statistics*, **19**, 51–65.

Hansen, M. H. and Hurwitz, W. N. (1943). On the theory of sampling from finite populations, *The Annals of Mathematical Statistics*, **14**, 333–362.

Rao, J. N. K. (1966). Alternative estimators in PPS sampling for multiple characteristics, *Sankhyā: The Indian Journal of Statistics, Series A*, **28**, 47–60.

Singh, S. (2003). *Advanced Sampling Theory with Applications: How Michael "Selected" Amy*, Kluwer Academic Publishers, Boston.

Warner, S. L. (1965). Randomized response: A survey technique for eliminating evasive answer bias, *Journal of the American Statistical Association*, **60**, 63–69.

Received July 2008; Accepted November 2008